Документация по Итоговой работе курса DWH

Описание базы данных для работы: <https://edu.postgrespro.ru/bookings.pdf>

Описание шагов работы:

1. Перед началом работы была изучена документация по бд. Выделены таблицы-источники данных для будущих операций.
2. Создано локальное подключение для дальнейшей работы. БД – postgres, схема – final.
3. Созданы таблицы с соответствующими названиями полей и форматами. Скрипты в приложении.  
   P.s. Таблица календарь создана на основе генератора дат, поэтому в переносе с бд не нуждается.   
   Чуть позже были созданы дополнительные таблички с пометкой reg для отсечения «кривых» записей по указанным условиям.
4. Было принято решение строить работу используя одну общую трансформацию. Сами трансформации находятся в приложении. Описание основных шагов представлено ниже.

Шаги трансформации:

1. Для таблицы fact\_flights
   1. Создаем инпут данных, с помощью подключения к бд нетологии.
   2. Также создаем отдельную переменную (системную) с сегодняшней датой (пригодится для отсева кривых дат)
   3. Джойним шаг a и b
   4. Следующим шагом выбираем необходимые колонки, правильно устанавливаем их порядок и формат
   5. Первое условие отсечения кривых строк – отсекаем случаи в данных, когда аэропорт прибытия и отлета совпадают. Если такая строка найдена – пишем данные в таблицу с пометкой reg, иначе идем дальше по шагам.
   6. Второе условие, отсекаем случаи, когда дата прилета раньше даты отлета. Если такая строка найдена, аналогично кидаем ее в reg. Иначе идем дальше.
   7. Третье условие, отсекаем случаи, когда дата отлета позже чем сегодня. Если такая строка найдена, аналогично кидаем ее в reg. Иначе идем дальше.
   8. Предпоследним шагом оставляем нужные колонки и форматируем.
   9. И последний шаг – запись в нашу локальную бд данных.
2. Для таблицы aircrafts
   1. Создаем инпут данных, с помощью подключения к бд нетологии.
   2. Следующим шагом выбираем необходимые колонки, правильно устанавливаем их порядок и формат
   3. Первое условие отсечения кривых строк – отсекаем случаи в данных, когда модель самолета равна нулю. Если такая строка найдена – пишем данные в таблицу с пометкой reg, иначе идем дальше по шагам.
   4. Второе условие, отсекаем случаи, когда дистанция меньше 100 км. Если такая строка найдена, аналогично кидаем ее в reg. Иначе идем дальше.
   5. Третье условие, отсекаем случаи, когда дистанция больше 40000 км. Если такая строка найдена, аналогично кидаем ее в reg. Иначе идем дальше.
   6. И последний шаг – запись в нашу локальную бд данных.
3. Для таблицы passenger
   1. Создаем инпут данных, с помощью подключения к бд нетологии.
   2. Следующим шагом выбираем необходимые колонки, правильно устанавливаем их порядок и формат
   3. Первое условие отсечения кривых строк – отсекаем случаи в данных, когда имя пассажира самолета пустое. Если такая строка найдена – пишем данные в таблицу с пометкой reg, иначе идем дальше по шагам.
   4. Второе условие, отсекаем случаи, когда контактные данные пассажира пустые. Если такая строка найдена, аналогично кидаем ее в reg. Иначе идем дальше.
   5. И последний шаг – запись в нашу локальную бд данных.
4. Для таблицы airports
   1. Создаем инпут данных, с помощью подключения к бд нетологии.
   2. Следующим шагом выбираем необходимые колонки, правильно устанавливаем их порядок и формат
   3. Первое условие отсечения кривых строк – отсекаем случаи в данных, когда наименование аэропорта равна нулю. Если такая строка найдена – пишем данные в таблицу с пометкой reg, иначе идем дальше по шагам.
   4. Второе условие, отсекаем случаи, когда город равен нулю. Если такая строка найдена, аналогично кидаем ее в reg. Иначе идем дальше.
   5. Третье условие, отсекаем случаи, когда таймзона равна нулю. Если такая строка найдена, аналогично кидаем ее в reg. Иначе идем дальше.
   6. И последний шаг – запись в нашу локальную бд данных.
5. Для таблицы tariff
   1. Создаем инпут данных, с помощью подключения к бд нетологии.
   2. Следующим шагом выбираем необходимые колонки, правильно устанавливаем их порядок и формат
   3. И последний шаг – запись в нашу локальную бд данных.